FAQ Retrieval using Query-Question Similarity and BERT-Based Query-Answer Relevance

Wataru Sakata†, Tomohide Shibata‡, Ribeka Tanaka‡, Sadao Kurohashi‡
†LINE Corporation, ‡Kyoto University

Issues
• Train the neural model using the relevance label between user’s q and appropriate QA [Karan+18]
  → It costs a lot to construct such labeled data.
• Using the relevance between a question and an answer trained by QA pairs in the community QA site such as Yahoo QA [Shafig+18] .
  → The size of QAs in FAQ is not often enough to train a model.

Proposed Retrieval Model

**Query : Can you please remove the wasp’s nest?**

**q-Q similarity**
• Due to the scarcity of the labeled data, we use the BM25-based unsupervised retrieval system, TSUBAKI [Shinzato+12]
• High precision but difficult to deal with lexical gap.

**q-A relevance**
• Trained with QA data using BERT [Devlin+19].
• Collecting other similar FAQ sets to increase the size of available QA data.
• It performs well even if there is a lexical gap between q and Q.

**Target FAQ**

Q: Please tell me how to take out the oversized garbage?
A: Please call the center, or you can also bring it there by yourself.

Q: I want to get a tax certificate by mail.
A: For the mail of the certificate, ...

Q: I want you to exterminate the bee’s nest.
A: When there are bee nests in residential buildings...

Q: I want to throw away large waste.

Combining TSUBAKI and BERT
1. For QA pairs whose q-Q similarity gets a higher score than α, we rank them in order of the similarity.
2. Then for others, we rank them in order of the weighted average of q-Q similarity and q-A relevance.

**How to use BERT?**

Q: I want to throw away a large trash?
A: Please call the center or you can also bring it there by yourself.

Q: Can I get a copy of my family register by mail?
A: Yes. To request the resident certificate with a personal number.

Q: Where is the location of the polling place?
A: There are three polling stations before the date in the city, ...

Q: Where and when can I get the resident certificate?
A: ... You can get the resident card on weekdays and Saturdays, ...

Search stage (for the city k)
q: I want to throw away a large waste.

Q1: Tell me how to take out the oversized trash.
Qb1: How to throw away oversized trash.
Qa1: Tell me how to take out the oversized garbage.
Qk1: Please tell me how to take out the oversized garbage.
Qk2: I want to get a tax certificate by mail.

Target Only FAQ set

**Experiments**

localgovFAQ dataset
• Collect 784 queries by crowdsourcing and 1786 QAs in FAQ page of Amagasaki City in Japan.
• Each QA pair was manually classified into four categories according to their relevance to query

Settings
• Use Japanese Wikipedia for pre-training.
• Collect 20k QA pairs, using FAQs of 21 Japanese local governments for the fine-tuning.
• Also evaluate in English using Stack Exchange dataset

**Hooks**

localgovFAQ(Japanese)

TSUBAKI (q-Q) 0.50
Bi-LSTM (q-A) 0.50
BERT targetOnly (q-A) 0.53
BERT (q-A) 0.56
COMBINED (q-QA) 0.62

Stack Exchange(English)

CNN-rank (q-Q) 0.79
TSUBAKI (q-Q) 0.70
BERT (q-A) 0.89
CNN-rank (q-QA) 0.74
COMBINED (q-QA) 0.90

**Discussions**

The accuracy of retrieved QA pair with high TSUBAKI score is very high.
Loose correlation between the accuracy and BERT score.

Correct Q: “Information on the facility of the youth center, hours of use, and closed day”

SIGIR ’2019 – Paris, France
Contact : wataru.sakata@linecorp.com